

Ответы на вопросы

Вопрос 1

Что такое LSTM блок? Для чего нужен?

Долгая краткосрочная память (Long short-term memory; LSTM) – особая разновидность архитектуры рекуррентных нейронных сетей, способная к обучению долговременным зависимостям. Используется для решения задач, в которых важно передать информацию от одного шага работы сети к другому (распознавание речи, перевод и т.д.). Чем больше разрыв между актуальной информацией и точкой ее применения, тем хуже обычные рекуррентные сети правятся со связыванием информации. LSTM разработаны специально, чтобы избежать проблемы долговременной зависимости.

Любая рекуррентная нейронная сеть имеет форму цепочки повторяющихся модулей нейронной сети. Структура LSTM также напоминает цепочку, но вместо одного слоя нейронной сети они содержат целых четыре, и эти слои взаимодействуют особым образом (это и есть LSTM блок).

Вопрос 2

В чем заключается смысл семантической сегментации, какого вида дает результат ИНС?

Семантическая сегментация изображений - это разделение изображения на отдельные группы пикселей, области, соответствующие одному объекту с одновременным определением типа объекта в каждой области.

ИНС дает результат в виде карты признаков.

Вопрос 3

За счет чего слой Dropout позволяет регулировать переобучение?

Главная идея Dropout — вместо обучения одной DNN обучить ансамбль нескольких DNN, а затем усреднить полученные результаты.

Сети для обучения получаются с помощью исключения из сети (dropping out) нейронов с вероятностью p , таким образом, вероятность того, что нейрон останется в сети, составляет $q = 1 - p$. "Исключение" нейрона означает, что при любых входных данных или параметрах он возвращает 0.

В стандартной нейронной сети производная, полученная каждым параметром, сообщает ему, как он должен измениться, чтобы, учитывая деятельность остальных блоков, минимизировать функцию конечных потерь. Поэтому блоки могут меняться, исправляя при этом ошибки других блоков. Это может привести к чрезмерной совместной адаптации (co-adaptation), что, в свою очередь, приводит к переобучению, поскольку эти совместные адаптации невозможно обобщить на данные, не участвовавшие в обучении. Dropout предотвращает совместную адаптацию для каждого скрытого блока, делая присутствие других скрытых блоков ненадежным. Поэтому скрытый блок не может

полагаться на другие блоки в исправлении собственных ошибок. Т.е. Dropout хорошо работает на практике, потому что предотвращает взаимонадаптацию нейронов на этапе обучения.

Вопрос 4

Обязательно ли форма ядра свертки должна быть квадратной?

Нет, допускается как прямоугольная, так и квадратная форма фильтра. Обычно используется квадратная размером 3x3, 5x5, 7x7 и 9x9.